
Apgar Scores: Examining the Long-term Significance

Kristen S. Montgomery, PhD, RNC, IBCLC

Abstract

The Apgar scoring system was intended as an evaluative measure of a newborn's condition at birth and of the need for immediate attention. In the most recent past, individuals have unsuccessfully attempted to link Apgar scores with long-term developmental outcomes. This practice is not appropriate, as the Apgar score is currently defined. Expectant parents need to be aware of the limitations of the Apgar score and its appropriate uses.

Journal of Perinatal Education, 9(3), 5–9; *Apgar scores, newborn, birth, resuscitation.*

Virginia Apgar, a physician and anesthesiologist, developed the Apgar scoring system in 1952 (Apgar, 1953) to evaluate a newborn's condition at birth. The Apgar score is performed at 1 and 5 minutes of life. The purpose of this paper is to discuss the appropriate use of the Apgar score and to examine the appropriateness of using the Apgar score to predict long-term developmental outcomes.

The Apgar scoring system is a comprehensive screening tool to evaluate a newborn's condition at birth (see Table 1). Newborn infants are evaluated based on five variables: heart rate, respiratory effort, muscle tone, reflex irritability, and color. A numerical score of 0–2 is assigned in each category for a maximum score of 10. Apgar scoring is best used in conjunction with additional evaluative techniques such as physical assessment and vital signs.

In recent years, many researchers have attempted to correlate Apgar scores with various outcomes including development (Behnke et al., 1989; Blackman, 1988;

KRISTEN S. MONTGOMERY is a recent doctoral graduate in nursing at Case Western Reserve University in Cleveland, Ohio. She resides in Clinton Township, Michigan.

Apgar Scores: Examining the Long-term Significance

Table 1 The Apgar Scoring System (from Apgar, V., 1966)

Sign	Score		
	0	1	2
Color	Pale Blue	Pink Body; Blue Extremities	Completely Pink
Reflex Irritability	None	Grimace	Vigorous Cry
Heart Rate	Absent	Slow (< 100)	Above 100
Respiratory Effort	Absent	Slow (irregular)	Crying
Muscle Tone	Flaccid	Some Flexion of Extremities	Active Motion
Score		Status	
7–10		Normal	
4–6		Moderately Depressed	
0–3		Severely Depressed	

Riehn, Petzold, Kuhlisch, & Distler, 1998), later delinquency (Gibson & Tibbetts, 1998), intelligence (Nelson & Ellenberg, 1981), and neurological development (Sommerfelt, Pedersen, Ellertsen, & Markestad, 1996; Wolf, M., Beunen, Casaer, & Wolf, B., 1998; Wolf, M., Beunen, Casaer, & Wolf, B., 1997; Wolf, M., Wolf, B., Bijleveld, Beunen, & Casaer, 1997) for the purposes of research. However, individuals have misinterpreted this research and, in some instances, attempted to apply causality (i.e., that low Apgar scores *caused* later delinquency or poor neurological outcomes). Causality has been neither established nor a goal of the currently reported research in this area. Research in this area has focused on establishing a *correlation* between these outcomes and an individual's Apgar scores. The aim was not to demonstrate that low Apgar scores caused or predicted these conditions; however, some individuals have incorrectly interpreted the research as stating low Apgar scores could predict or actually caused certain behaviors or deficits. Not only is this inappropriate use of the Apgar score, there is also little scientific evidence

Apgar scoring is best used in conjunction with additional evaluative techniques such as physical assessment and vital signs.

to support its use in predicting long-term outcomes. Please see Table 2 for clarification of selected research terms.

. . . there is also little scientific evidence to support [the Apgar score's] use in predicting long-term outcomes.

Review of the Literature

According to the American Academy of Pediatrics' Committee on Fetus and Newborn and the American College of Obstetricians and Gynecologists' Committee on Obstetric Practice (1996), the Apgar score should be used to assess the condition of an infant at birth. These committees also warn that the Apgar score should not be used as the only measure to evaluate the possibility that neurological damage occurred during the birthing process. In addition to low Apgar scores (3 or less for longer

Table 2 Glossary of Selected Research Terms

Causal Relationship

A relationship between two variables such that the presence or absence of one variable (the "cause") determines the presence or absence, or value, of the other (the "effect").

Correlation

A tendency for variation in one variable to be related to variation in another variable.

Empirical Evidence

Evidence that is rooted in objective reality and gathered through the collection of data using one's senses; used as the basis for generating knowledge through the scientific approach.

Population

The entire set of individuals (or objects) having some common characteristic(s) (e.g., all RNs in the state of California); sometimes referred to as a universe.

Prediction

One of the aims of the scientific approach; the use of empirical evidence to make forecasts about how variables of interest will behave in a new setting and with different individuals.

Relationship

A bond or a connection between two or more variables.

Variable

A characteristic or attribute of a person or object that varies (i.e., takes on different values) within the population under study (e.g., body temperature, age, heart rate).

than 5 minutes), an infant who is asphyxiated prior to delivery would demonstrate severe metabolic or mixed acidemia (pH < 7.00) via umbilical artery blood sample and additional neurological manifestations such as seizure activity, coma, hypotonia, and finally, evidence of multiorgan dysfunction (Committee on Fetus and Newborn, American Academy of Pediatrics, & Committee on Obstetric Practice, American College of Obstetricians and Gynecologists, 1996). Furthermore, research conducted over 30 years ago (Apgar, 1966; Apgar & James, 1962) provided initial evidence to disclaim the reliability of Apgar scores for predicting long-term outcomes of any type (e.g., developmental and neurological). Prediction of long-term outcomes was never a goal of the Apgar scoring system. Rather, the goal was to make certain that infants were systematically observed for their need for immediate care at birth.

Reliability of Apgar Scores

According to Jepson, Talashek, and Tichy (1991), the Apgar score as a “tool” (to measure newborn adaptation to extrauterine life) lacks sensitivity and specificity. *Sensitivity* measures how well the tool captures the infant’s condition at birth (stable vs. depressed) and *specificity* refers to how well the tool measures the differences between the values of the scores (0–2 for each of the five categories). Additionally, various authors have noted that great variability exists in how individual health care providers score the assessment (Clark & Hakanson, 1988; Livingston, 1990). Clark and Hakanson (1988) compared the consistency (inner-rater reliability) of Apgar scoring among various health care disciplines. In their study, groups of health care providers were visually shown case presentations and then asked to assign Apgar scores to the infants who were presented. Pediatricians and pediatric house staff had a consistency rating of 68%, obstetricians and obstetric house staff had a consistency rating of 46%, intensive care nursery staff had a consistency score of 42%, obstetric nurses 36%, and community hospital nurses a consistency rating of 24%.

Livingston examined how consistent two health care providers were in assigning scores when compared to one another. In this study, the consistency of scores ranged from 55% to 82% with heart rate having the best rate of consistency at 82% for the 1-minute scores (Livingston, 1990). For the 5-minute score, consistency

ranged from 36% to 100%, again heart rate having the highest rate of consistency. Heart rate measures likely have greater consistency due to the ease of understanding and defining exactly what is being assessed. When consistency scoring was compared between full-term and premature newborns, health care providers were found to have better consistency when assessing full-term newborns (Livingston, 1990). Additionally, full-term newborns may represent the “normal” in health care provider’s minds; hence, full-term newborns may be more likely to receive a “normal” score, which accounts for the higher rate of consistency in term newborns than in preterm newborns.

Another concern is determining who has responsibility for assigning the Apgar score once the infant is born. According to both Apgar (1966) and the Regan Report (1987), the person assisting with the delivery of the infant should not assign the Apgar score. While in some respects the delivering individual seems the most logical choice, bias may be introduced into the score value, because the individual who attends the delivery may have a vested interest in the outcome.

Secondly, the newborn may be given to additional personnel immediately after delivery. This makes determining the Apgar score considerably more difficult for the health care provider who is assisting the delivery, necessitating leaving the mother’s bedside briefly to assign the score. Additionally, if the infant remains with the mother for the first 5 minutes of life, the health care provider must later remember to document the score, often from memory. Both circumstances have the potential to introduce further bias to the already poor consistency of the Apgar score.

Often the nurse or someone from the department of neonatology assigns the Apgar score. Most frequently in a normal, full-term delivery, this would be the nurse. Nurses, at least in the Clark and Hakanson (1988) study, had a poor consistency rate. Questions regarding the accuracy of the Apgar score play a role in limiting the long-term predictive value.

Intended Uses of the Apgar Score

As the Apgar score was developed and refined over the years since its inception, the intended use has always been the same: to evaluate a newborn’s condition at birth. Some clinicians like to use the Apgar score as a

guide to their resuscitative efforts; however, this is not an intended use of the Apgar score. The novice practitioner may mistakenly believe that resuscitative efforts should not begin until the 5-minute Apgar score is determined. Experienced clinicians realize this would severely delay resuscitative efforts and compromise the potential for full recovery of neurological function. It is important to be both careful and consistent with language.

Educating the Public

Letko (1996) notes that much of the public, especially expectant parents, has some level of familiarity with the Apgar score. However, as Letko also points out, many of these parents-to-be do not adequately understand the score or its capacities for predicting long-term outcomes. Parents need to receive the appropriate education through the popular media, childbirth classes, and health care providers. It is imperative that parents have appropriate information so they are not disappointed when their child receives a score of 9, believing that their child is somehow inadequate because he or she did not receive a score of 10. Parents need to understand that a score from 7–10 indicates a normal newborn at birth and that it is rather infrequent for a newborn to receive a score of 10. For example, most infants have some level of blueness to their extremities and will not initially be completely pink. This point can be covered when discussing the general appearance of a newborn. This anticipatory guidance can assist the parents in their understanding and promote a positive birthing experience, by avoiding potential disappointment.

It is imperative that parents have appropriate information so they are not disappointed when their child receives a score of 9, believing that their child is somehow inadequate because he or she did not receive a score of 10.

Summary

The Apgar scoring system was intended as an evaluative measure of a newborn's condition at birth and of the

need for immediate attention. In the most recent past, individuals have unsuccessfully attempted to link Apgar scores with long-term developmental outcomes. This practice is not appropriate as the Apgar score is currently defined. Expectant parents need to be aware of the limitations of the Apgar score and its appropriate uses.

Directions for Future Research

Future research is needed to increase the consistency among health care providers assigning Apgar scores. This would take a training program and periodic practice sessions to establish and maintain inner-rater reliability of each professional whose role is to assign the scores. Enhanced consistency would be a first step to evaluating the effectiveness of Apgar scores. At present, Apgar scores serve as a somewhat useful screening tool for health care providers to communicate with each other about what a newborn's status was like at birth and as a mechanism to make certain that someone is systematically observing the condition of the new infant.

Acknowledgement

The author thanks Raquel Mayne, BSN, graduate of New York University Division of Nursing, for her assistance with article retrieval.

References

- Anonymous. (1996) Use and abuse of the Apgar score. Committee on Fetus and Newborn, American Academy of Pediatrics, and Committee on Obstetric Practice, American College of Obstetricians and Gynecologists. *Pediatrics*, 1, 141–142.
- Apgar, V. (1953). A proposal for a new method of evaluation of the newborn infant. *Current Researches in Anesthesia and Analgesia*, 32, 260–267.
- Apgar, V. (1966). The newborn scoring system: Reflections and advice. *Pediatric Clinics of North America*, 113, 645–650.
- Apgar, V., & James, L. (1962). Further observation of the newborn scoring system. *American Journal of Diseases of Children*, 104, 419–428.
- Behnke, M., Eyer, F., Carter, R., Hardt, N., Cruz, A., & Resnick, M. (1989). Predictive value of Apgar scores for developmental outcome in premature infants. *American Journal of Perinatology*, 6, 18–21.
- Blackman, J. (1988). The value of Apgar scores in predicting

-
- developmental outcome at age five. *Journal of Perinatology*, 8, 206–210.
- Clark, D., & Hakanson, D. (1988). The inaccuracy of Apgar scoring. *Journal of Perinatology*, 8, 203–205.
- Gibson, C., & Tibbetts, S. (1998). Interaction between maternal cigarette smoking and Apgar scores in predicting offending behavior. *Psychological Reports*, 83, 579–586.
- Jepson, H., Talashek, M., & Tichy, A. (1991). The Apgar score: Evolution, limitations, and scoring guidelines. *Birth: Issues in Perinatal Care*, 18, 83–92.
- Letko, M. (1996). Understanding the Apgar score. *Journal of Obstetrical, Gynecological, and Neonatal Nursing*, 25, 299–303.
- Livingston, J. (1990). Interrater reliability of the Apgar score in term and premature infants. *Applied Nursing Research*, 3, 164–165.
- Nelson, K., & Ellenberg, J. (1981). Apgar scores as predictors of chronic neurologic disability. *Pediatrics*, 68, 36–44.
- Riehn, A., Petzold, C., Kuhlisch, E., & Distler, W. (1998). Fetal acidemia and neonatal encephalopathy. *Zeitschrift für Geburtshilfe und Neonatologie*, 202, 187–191. (abstract).
- Sommerfelt, K., Pedersen, S., Ellertsen, B., & Markestad, T. (1996). Transient dystonia in non-handicapped low-birthweight infants and later neurodevelopment. *Acta Paediatrica*, 85, 1445–9.
- _____. Watch those Apgar scores: Evidence. (1987). *Regan Report on Nursing Law*, 27, 4.
- Wolf, M., Beunen, G., Casaer, P., & Wolf, B. (1998). Neonatal neurological examination as a predictor of neuromotor outcome at 4 months in term low-Apgar-score babies in Zimbabwe. *Early Human Development*, 51, 179–186.
- Wolf, M., Beunen, G., Casaer, P., & Wolf, B. (1997). Neurological findings in neonates with low Apgar in Zimbabwe. *European Journal of Obstetrics, Gynecology, & Reproductive Biology*, 73, 115–119.
- Wolf, M., Wolf, B., Bijleveld, C., Beunen, G., & Casaer, P. (1997). Neurodevelopmental outcome in babies with a low Apgar score from Zimbabwe. *Developmental Medicine & Child Neurology*, 39, 821–826.